wps 1553

POLICY RESEARCH WORKING PAPER 1553

Spatial Correlations in Panel Data

A correction for spatial correlation in panel data.

John Driscoll
Aart Kraay

The World Bank
Policy Research Department
Macroeconomics and Growth Division
December 1995

Summary findings

In many empirical applications involving combined time-series and cross-sectional data, the residuals from different cross-sectional units are likely to be correlated with one another. This is often the case in applications in macroeconomics and international economics where the cross-sectional units may be countries, states, or regions observed over time. "Spatial" correlations among such cross-sections may arise for a number of reasons, ranging from observed common shocks such as terms of trade or oil shocks, to unobserved "contagion" or "neighborhood" effects which propagate across countries in complex ways.

Driscoll and Kraay observe that the presence of such spatial correlations in residuals complicates standard inference procedures that combine time-series and cross-sectional data since these techniques typically require the assumption that the cross-sectional units are independent. When this assumption is violated, estimates of standard errors are inconsistent, and hence are not useful for inference. And standard corrections for spatial correlations will be valid only if spatial correlations are of particular restrictive forms.

Driscoll and Kraay propose a correction for spatial correlations that does not require strong assumptions concerning their form — and show that it is superior to a number of commonly used alternatives.

This paper — a product of the Macroeconomics and Growth Division, Policy Research Department — is part of a larger effort in the department to study international macroeconomics. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Rebecca Martin, room N11-059, telephone 202-473-9065, fax 202-522-3518, Internet address rmartin1@worldbank.org. December 1995. (28 pages)

Spatial Correlations in Panel Data [1]

John Driscoll
Department of Economics
Brown University, Box B
Providence RI 02912
jd@econ.pstc.brown.edu

and

Aart Kraay
The World Bank
1818 H Street NW
Washington DC 20433
akraay@worldbank.org

[1] We would like to thank John Campbell, Greg Mankiw, Matthew Shapiro and especially Gary Chamberlain and Jim Stock for helpful comments and suggestions. Financial support from the Earle A. Chiles Foundation (Driscoll) and the Social Sciences and Humanities Research Council of Canada (Kraay) during work on earlier drafts of this paper is gratefully acknowledged.

1 Introduction

Economists are frequently faced with the problem of drawing inferences from data sets which combine cross-sectional and time-series data. In such situations, it has become standard practice to base inferences on techniques which pool the cross-sectional and time-series dimensions in some way. For such techniques to be valid, it must be the case that the error terms are not correlated across different cross-sectional units, either contemporaneously or at leads and lags. This condition is directly analogous to the usual requirement that the residuals from different observations in a single cross-sectional regression be independent of each other. If this condition is not met, estimates of standard errors will be inconsistent, and will not be useful for inference.

This paper begins with the observation that in many applications, especially in macroeconomics and international economics, the assumption of independent cross-sectional units is inappropriate. While it may be reasonable to assume that cross-sectional units are independent when they are households or individuals chosen according to a well-designed sampling scheme from a large population, this assumption becomes less tenable when the cross-sectional units are countries or regions. Countries or regions are likely to be subject to observable and unobservable common disturbances which will cause the residuals from one cross-section to be correlated with those of another. We will refer to such cross-sectional correlations as "spatial correlations"

Spatial correlations may arise for a number of reasons. For example, in applications in which real GDP growth rates are the dependent variable, various channels of interdependence such as trade, capital flows or policy coordination mechanisms will induce cross-country correlations in GDP growth rates.[1] Unless the regressions of interest include right-hand side variables which correctly specify these channels of interdependence, the residuals from these regressions will be correlated across countries. Similarly, in studies of capital flows to developing countries, common external shocks such as US interest rates, or else unobserved contagion effects

[1] See Kraay and Ventura (1995) for a discussion of the roles of trade and capital mobility in the synchronization of GDP growth rates across countries. Ades and Chua (1993) and Easterly and Levine (1995) provide empirical evidence that policies tend to be correlated among neighbours, leading to correlations of growth rates over long horizons.

(sometimes dubbed "tequila" effects in aftermath of the Mexican peso crisis) can cause residuals from capital flows regressions to be correlated across countries.

A number of standard corrections for spatial correlations exist, all of which require strong assumptions regarding the form of the spatial correlations. For example, it is common to include time dummy variables in pooled time-series, cross-sectional regressions to capture the effect of common disturbances. This technique is the appropriate correction for spatial correlation only if one assumes that the contemporaneous correlations between any pair of cross-sectional units are equal, and the lagged cross-sectional correlations are zero. Unfortunately, such strong restrictions on the form of the spatial correlations are unlikely to be correct in most applications. For example, different countries may react differently to common disturbances, or contagion effects may spread across countries only after a lag. When the structure of the spatial correlations is misspecified in this way, the properties of the resulting estimator are in general unknown.

Since it is not desirable to impose restrictions on the form of the spatial correlations, it is less clear how to proceed. One alternative is to attempt to parametrically estimate the full unrestricted matrix of spatial correlations for use in a feasible generalized least squares (FGLS) procedure. This procedure, which is a variant of the Seemingly Unrelated Regressions (SUR) technique, will only be effective in a limited set of applications. To see why this is so, suppose that there are N cross-sectional units and T time-series observations. The NxN matrix of contemporaneous cross-sectional correlations has $N(N+1)/2$ free parameters to be estimated using the NT available observations. Thus, in order to obtain reliable estimates of the matrix of spatial correlations, it must be the case that $T>>(N+1)/2$. However, in many cross-country applications using annual data, there are many more countries in the sample of interest than there are time-series observations, so this approach will be infeasible.

In this paper we propose an alternative correction for spatial correlation. Building on the non-parametric heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation technique of Newey and West (1987) and Andrews (1991), we show how this approach can be extended to a panel setting with cross-sectional dependence, in addition to serial correlation and heteroskedasticity. We present very weak conditions on the form of the cross-sectional and time-series dependence under which a simple variant on the Newey and West estimator yields consistent estimates of standard errors. In particular, we can obtain consistent estimates of standard errors in the presence of arbitrary contemporaneous cross-sectional correlations, as well as lagged cross-sectional correlations which are restricted to become small only as the time interval separating the two observations becomes large. This very general structure is likely to encompass most forms of spatial correlations encountered in practice.

Our results on consistency are based on asymptotic theory which requires the time dimension, T, to tend to infinity. Thus, our results will only be relevant for panel data sets in which the time dimension is reasonably large (our Monte Carlo simulations suggest that a value of T=20 or T=25 is the minimum). However, our results do not place any restrictions on the size of the cross-sectional dimension, N, and we can even allow the extreme case in which N tends to infinity at any rate relative to T. This implies that our techniques, in contrast to SUR, will be applicable in situations such as cross-country panel data sets where the number of countries is very large.

The rest of this paper proceeds as follows. In Section 2, we first develop the intuitions for our results using a simple ordinary least squares example. We then provide a formal statement of our result, using a mixing random field structure to characterize the permissible extent of cross-sectional and time-series dependence. Since this structure is somewhat unfamiliar, we provide some examples of forms of cross-sectional dependence which satisfy the conditions we impose. In Section 3, we consider the finite-sample properties of our estimator using Monte Carlo evidence, and find that our non-parametric estimator performs significantly better than common alternatives such as time dummies or SUR. Section 4 concludes.

2 Consistent Covariance Matrix Estimation with Spatial Dependence

2.1 Preliminary Discussion

In order to develop the intuition for the results of this paper, consider the following simple bivariate linear panel regression:

$$\begin{gathered} y_{it} = x_{it}\beta + \epsilon_{it} \\ i = 1,\ldots,N, \quad t = 1,\ldots,T \\ \left\{ E[\epsilon_{it}\epsilon_{js}] \right\} = \Omega \end{gathered} \tag{1}$$

To obtain an estimate of β, it is common practice to pool the cross-sectional and time-series observations and apply OLS to the full set of NT observations. If the errors are independently and identically distributed (i.e. if $\Omega = \sigma^2 I_{NT}$), this will yield consistent estimates of β and its standard errors. However, in the presence of spatial correlations, Ω is no longer diagonal. In this case, although the OLS estimator of β is still consistent, the OLS standard errors will be inconsistent, and hence will not be useful for inference.

We can write the OLS estimator of β in the usual way as follows:

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) = \frac{\sqrt{T}\sum_{t=1}^{T}\sum_{i=1}^{N} x_{it}\epsilon_{it}}{\left\{ \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} x_{it}^2 \right\} NT} \tag{2}$$

To simplify the above expression, denote the term in brackets in the denominator of **(2)** as Q_T [2], and define

$$h_t \equiv \frac{1}{N}\sum_{i=1}^{N} x_{it}\epsilon_{it} \tag{3}$$

---

[2] For the purposes of this illustrative example, we can assume that the x_{it} are constants and that $Q_T \to Q > 0$ as $N,T \to \infty$.

Substituting into the expression for the OLS estimator, we obtain

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) = \frac{1}{Q_T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t \tag{4}$$

This change of variables is useful because it reduces the original panel data estimation problem to a simple time-series estimation problem. In other words, by defining a cross-sectional average h_t at every point in time, we have "collapsed" the cross-sectional dimension of the problem to a single time-series observation by averaging over the N cross-sectional units in each period.

Since OLS estimates of β will be consistent even in the presence of spatial correlations, our main concern is with obtaining consistent estimates of the variance of the OLS estimator. Using the above notation, we can write this variance in terms of the h_t as

$$V_T = \frac{1}{Q_T^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E[h_t h_s] = \frac{S_T}{Q_T^2} \tag{5}$$

The main intuition of the paper is as follows. Given appropriate conditions on the h_t, we can apply standard time-series non-parametric covariance matrix estimation techniques such as those employed by Newey and West to obtain a consistent estimate of S_T, and hence of V_T. These conditions (known as "mixing conditions" in the standard time-series literature) place restrictions on the autocovariances of the h_t, requiring the dependence between h_t and h_{t-s} to become small as the time interval separating them, s, becomes large. Imposing restrictions on the autocovariances of h_t will amount to placing restrictions on the contemporaneous and lagged spatial dependence in the residuals, $E[\epsilon_{it}\epsilon_{j,t-s}]$, since the autocovariances of the sequence h_t are a weighted average of these covariances, i.e.

$$E[h_t h_{t-s}] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{it}x_{j,t-s}E[\epsilon_{it}\epsilon_{j,t-s}] \tag{6}$$

In this paper, we show that only very weak restrictions on the form of the spatial correlations are required to ensure that h_t satisfies the regularity conditions necessary for consistent estimation of S_T. In particular, we can permit arbitrary contemporaneous correlations, and we require only that lagged cross-sectional dependence declines at a particular rate as the time separation becomes large. As in Newey and West (1987), our asymptotic results rely on a large time dimension. However, we do not need to restrict the size of the cross-sectional dimension, which can tend to infinity at any rate relative to T.

We use a mixing random field structure to characterize the permissible extent of spatial and temporal dependence. As mixing random fields are somewhat unfamiliar in the econometrics literature, we briefly present the necessary intuitions here, and relegate the details to the appendix. Random fields are simply random variables with multiple indices. For example, returning to Equation **(3)**, we can define the random field $h_{it} \equiv x_{it}\epsilon_{it}$, indexed by i and t. In the standard univariate time-series literature, a time series is described as "mixing" if the dependence between two random variables x_t and x_{t-s} becomes small as the time interval separating them, s, becomes large. In this paper, we will analogously describe a random field as being "mixing" if the dependence between h_{it} and $h_{j,t-s}$ becomes small as the time interval s becomes large, for any pair of cross-sectional observations i and j.[3] In this way, the standard time-series definition of mixing corresponds to the special case where i=j. Finally, the "size" of a mixing is defined as the rate at which the dependence between two observations must decline as a function of the distance between them.

This particular definition of a mixing random field has the extremely useful property that the cross-sectional averges of this random field, h_t (as defined in Equation **(3)**), form a univariate

[3] This definition of mixing departs from the standard definitions in the random field literature in that it treats the cross-sectional and time-series dimensions asymmetrically. Typically, mixing restriction would require the dependence between h_{it} and $h_{j,t-s}$ to become small as the Euclidean distance $d=((i-j)^2+s^2)^{1/2}$ between these two random variables becomes large. This is an unattractive property of standard definitions of mixing random fields for two reasons. First, in our panel data applications, it precludes canonical forms of cross-sectional dependence such as equal contemporaneous cross-unit correlations. To see why this is so, notice that the distance between h_{it} and h_{jt} is simply $|i-j|$ according to the above definition. Standard definitions of mixing would then rule out equal cross-sectional correlations between any h_{it} and h_{jt}, since this correlation will not decline as $|i-j|$ becomes large. The second problem is that in order to impose the restriction that observations "far apart" in the cross-sectional ordering be approximately uncorrelated, it is necessary to know what the cross-sectional ordering is. This is problematic, since unlike in the time dimension, in most cases there is no natural ordering in the cross-sectional dimension.

mixing sequence of the same size as the underlying random field. This is true for any value of N (the size of the cross-sectional dimension), including the limiting case where $N \to \infty$. If we impose the restriction that the h_{it} form a mixing random field of the appropriate size, then h_t will be a mixing sequence of the same size, and we can directly apply standard time-series covariance matrix estimation techniques to obtain an estimate of S_T in Equation (5). Thus, our results amount to a simple extension of the Newey and West estimator, which may be viewed in the above context as the case in which N=1.

2.2 Results

In this section, we present our main result, which is simply a generalization and formalization of the discussion of the previous section. The theorem is stated in terms of a broad class of Generalized Method of Moments estimators, of which the OLS case discussed above is an example.

Theorem

Consider the class of GMM models identified by a px1 vector of orthogonality conditions $E[\psi(\theta_0, z_{it})]=0$, *where* $\theta_0 \in \Theta$ *is an* ax1 *vector of parameters with* $a \leq p$, Θ *is a compact subset of* R^a, z_{it} *is a kx1 vector of data, and denote* $z_t=(z_{it}',\ldots,z_{Nt}')'$ *and* $h_t=h(\theta, z_t)=N^{-1}\Sigma_{i=1}^{N}\psi(\theta, z_t)$. *Suppose further that*

(1) z_{it} *is an* α-*mixing random field of size* $2(r+\delta)/(r+\delta-1)$, *as defined in the Appendix;*

(2) (a) $\psi(\theta, z_{it})$ *is continuously differentiable in* θ *and measurable in* z_{it};

(b) $E[|\psi(\theta, z_{it})|^{4(r+\delta)}]<\Delta<\infty\ \forall\theta\in\Theta$;

(3) *For all sequences* $\{\theta_T^*\}$ *such that* θ_T^* *converges in probability to* θ_0, $plim\{T^{1}(\partial h(\theta, z_t)/\partial\theta')|\theta=\theta_T^*\}=plim\{T^{1}(\partial h(\theta, z_t)/\partial\theta')|\theta=\theta_0\}=D'$ *where* D' *is of full column rank.*

Then the GMM estimator

$$\hat{\theta}_T = argmin_{<\theta>}\left[\frac{1}{T}\sum_{t=1}^{T}h_t(\theta,z_t)\right]'\hat{S}_T^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}h_t(\theta,z_t)\right]$$

is consistent and asymptotically normal and the panel Newey and West (1987) covariance matrix estimator

$$\hat{V}_T = \hat{D}_T^{-1}\left(\frac{1}{T}\sum_{j=-m(T)}^{m(T)}w(j,m)\sum_{1<t,t-j<T}h(\hat{\theta},z_t)h(\hat{\theta},z_{t-j})'\right)\hat{D}_T^{-1'} = \hat{D}_T^{-1}\hat{S}_T\hat{D}_T^{-1'}$$

is a consistent estimator of V_T *as* $T\to\infty$ *for any N (including* $N\to\infty$*), where* $w(j,m)=1-j/(m(T)+1)$, $m(T)=O(T^{1/3})$ *and* $\hat{D}_T$ *is a consistent estimator of* D_T.

Proof: See Appendix.

Before presenting some examples of the forms of spatial correlations encompassed by the theorem, a few comments are in order. First, the random field structure in Assumption (1) is the only substantive assumption required for the above result, as the remaining assumptions are fairly standard conditions required to establish the properties of the GMM estimator. Note that the assumptions of the theorem require no prior knowledge of the form of the spatial and temporal correlations, and place only very weak restrictions on them. Hence this framework subsumes many common forms of spatial dependence, without requiring an explicit (and probably also incorrect) parameterization of the form of the temporal and spatial dependence.

Next, a sketch of the proof is as follows. The regularity conditions placed on $\psi(\theta, z_{it})$ in Assumption 2(a) are sufficient to ensure that ψ itself is a mixing random field of the same size as z_{it}. The cross-sectional averages of this random field, h_t, will form a univariate mixing sequence as described above. The remainder of the proof is then simply a matter of verifying the standard results for the consistency and asymptotic normality of the GMM estimator and the consistency of the Newey and West covariance matrix estimator.

Finally, as an example of an application of the theorem, consider the simple OLS example of the previous section, which is a particular case of the GMM framework in which the above theorem is cast. To see this, set $\theta_0=\beta$, $z_{it}=(y_{it}, x_{it})$ and $\psi(\beta,x_{it})=x_{it}(y_{it}-\beta x_{it})$. Thus to implement this technique, first obtain the usual OLS coefficient estimate. Next construct the sequence of cross-sectional averages h_t using the estimated residuals from the OLS specification (i.e. $h_t=N^{-1}\sum_{i=1}^{N}x_{it}(y_{it}-bx_{it})$, where b is the OLS estimate of β) and insert this into the definition of S_T to obtain its consistent estimate. Finally, note that a consistent estimate of D_T is given by $(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}x_{it}^2$. Combining these two expressions gives the consistent estimate of the covariance matrix, V_T.[4, 5]

---

[4] TSP and GAUSS codes to perform these calculations are available from the authors.

[5] Another commonly-encountered model is one with unit-specific fixed effects, $y_{it}=f_i+x_{it}\beta+\epsilon_{it}$. One method to apply the theorem to this model is to transform the data by taking deviations from unit means and rewriting as above i.e. letting $z_{it}=(y_{it}-T^{-1}\sum_{t=1}^{T}y_{it}, x_{it}-T^{-1}\sum_{t=1}^{T}x_{it})$.

2.3 Examples of Spatial Correlation

As the mixing random field assumption of the theorem may be somewhat difficult to verify in practice, in this section we present some simple examples of forms of spatial correlation which satisfy this assumption. It is most convenient to present some examples of permissible forms of spatial correlation using the simple linear model with fixed scalar regressors of Equation **(1)**. A broad class of spatial correlations can be represented using the following factor structure[6] for the residuals of this regression, i.e.

$$\epsilon_{it} = f_t'\lambda_i + v_{it} \tag{9}$$

f_t is an Mx1 vector of independent, zero mean, unit variance random variables referred to as "factors", while λ_i is an Mx1 vector of scalar "factor loadings". The v_{it} are random variables with zero mean and variance σ_{it}^2 which are independent over time and across units. In addition, they are orthogonal to the factors, and are referred to as the "residuals from the factor structure". In the appendix, we show that this factor structure satisfies the conditions of the theorem.[7] This result thus provides a convenient way to verify the somewhat more abstract mixing random field conditions of the theorem, and describes a broad class of spatial correlations which have a factor structure representation to which our covariance matrix estimator is robust.

The simplest case of spatial correlations is one in which the contemporaneous cross-sectional correlations are all equal. It is straightforward to verify that if M=1, $\lambda_i=\lambda$ for all i, and $\sigma_{it}^2=\lambda^2-\sigma^2$ for all i and t, then the OLS residuals ϵ_{it} have mean zero, variance σ^2 and cross-sectional correlations $E[\epsilon_{it}\epsilon_{jt}]=\lambda^2$. In this special case, including time dummies in Equation **(1)** will remove all the spatial correlations from the residuals.

A more interesting case of spatial correlations with a factor structure representation is one

---

[6] We are grateful to Gary Chamberlain for suggesting this approach. See Chamberlain and Rothschild (1983) for a discussion of factor structures in the context of finance theory, and Al-Najjar (1995) for a discussion of factor structures as a method of modelling aggregate uncertainty in a continuum of cross-sectional units.

[7] The result in the appendix does not rely on the assumption that the residuals from the factor structure are contemporaneously uncorrelated across units, but allows them to have arbitrary spatial correlations.

in which the cross-sectional units are divided into m=1,...,M groups, and the within-group correlations are equal for all the members of the group. Such groups might be regions, geographic "neighbours" or any other grouping based on observable or unobservable characteristics. To represent this as a factor structure, define I_m as a set which consists of the indices of the members of group m. Then, if $\lambda_{mi}=\lambda_m$ for $i\in I_m$, $\sigma_{it}^2=\lambda_{mi}^2-\sigma^2$, and the v_{it} are again independent across units, it is immediate to verify that the OLS residuals will have mean zero, variance σ^2 and cross-sectional correlations $E[\epsilon_{it}\epsilon_{jt}]=\lambda_m^2$ for $i,j\in I_m$ and zero otherwise.

The most general case of contemporaneous spatial correlations is one in which the cross-sectional correlations are arbitrary. This would be a natural structure in the case where there is a common factor to which cross-sectional units react differently. If we introduce the assumption that the residuals from the factor structure, v_{it} have arbitrary contemporaneous cross-sectional correlations, then we can somewhat trivially write this as a factor structure in which all the factor loadings are zero. In this case, there will be spatial correlations in the residuals from Equation (1) even after time dummies are included in the specification.

Note that in these examples, we have used a factor structure to characterize the contemporaneous spatial dependence in the residuals. We can easily extend this to introduce dependence over time as well. For example, suppose that there are arbitrary contemporaneous cross-sectional correlations, zero lagged cross-sectional correlations, and within units the disturbances follow an AR(M) process. To give this a factor structure representation, let f_m have an autocovariance function which is 1 at lag m, and zero otherwise. Then if we set $\lambda_{mi}=E[\epsilon_{it}\epsilon_{i,t-s}]$, the factor structure will replicate exactly this combination of cross-sectional and temporal dependence. Along the same lines, it is possible to generate much more complicated forms of lagged cross-sectional dependence using this factor structure.

3 Monte Carlo Results

In this section we use Monte Carlo experiments to examine how well our estimator (which we will refer to as the HAC estimator) performs relative to common alternative corrections for spatial correlations such as SUR and OLS and time dummies.[8] We generate large numbers of samples of artificial data with various forms of spatial correlations, and obtain coefficient estimates and standard errors using OLS, SUR and our HAC estimator. We can then evaluate the relative performance of these estimators by reporting a "coverage rate" for each estimator, which is the fraction of samples in which two standard deviation confidence intervals contain the true parameter values. For the HAC procedure, this fraction equals .95 as $T \to \infty$, as it does not rely on a correct parameterization of the temporal and spatial correlations. In finite samples ($T<\infty$), however, this coverage rate will inevitably be somewhat smaller. OLS with time dummies and SUR will in general be misspecified, and their asymptotic properties when misspecified are generally unknown. However, by reporting coverage rates, we can get a rough idea of the severity of the impact of the misspecification of these procedures.

We consider linear models such as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ . \\ . \\ y_{Nt} \end{pmatrix} = \begin{pmatrix} x_{1t} \\ x_{2t} \\ . \\ . \\ x_{Nt} \end{pmatrix} \beta + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ . \\ . \\ \epsilon_{Nt} \end{pmatrix}, \quad t=1,\ldots,T \tag{10}$$

Without loss of generality, we set $\beta=0$ in Equation **(10)**. To introduce a rich structure of

---

[8] Most other techniques are versions of feasible GLS which impose various zero restrictions on the variance-covariance matrix (for example, Case (1991) and Keane and Runkle (1992)). Elliot (1993) discusses the case when there is a single cross-section, with one observation per geographical location. There are two classes of interesting exceptions to this. One class, used in finance, assumes there is no serial correlation and applies a technique similar to the White (1980) correction for heteroskedasticity (for example, Fama and MacBeth (1973), Lehmann (1990), and Froot (1989)). A second alternative has been offered by Conley (1994), who proposes a nonparametric variance-covariance matrix estimator when the "distance" between cross-sectional units is known.

contemporaneous and lagged spatial correlations into the residuals, we generate them according to the following autoregressive scheme:

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ . \\ . \\ \epsilon_{Nt} \end{pmatrix} = R \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ . \\ . \\ \epsilon_{N,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{\epsilon_{1t}} \\ \eta_{\epsilon_{2t}} \\ . \\ . \\ \eta_{\epsilon_{Nt}} \end{pmatrix} \quad , \quad R = \rho I_N \quad , \quad \{E[\eta_{\epsilon_{it}} \eta_{\epsilon_{jt}}]\} = \Sigma \tag{11}$$

This specification ensures that the residuals in Equation **(10)** exhibit both contemporaneous and lagged spatial correlations. In the simple case in which $\rho=0$, there are only contemporaneous spatial correlations in the residuals in Equation **(10)**, and these are given by Σ. In this case, both the HAC and the SUR are correctly specified, and their coverage rates can be compared directly. Finally, we use the same structure to generate the regressors, x_{it}.

Before we can perform Monte Carlo experiments, we need to parameterize the spatial correlation matrix Σ and the serial correlation parameter, ρ. We allow the parameter ρ to range over the values {0, .1, .3, .5}, which correspond to the moderate degree of serial correlation likely to be present in most applications. Selecting the matrix of contemporaneous spatial correlations is more difficult. One alternative is to choose a simple parameterization for this matrix, and report Monte Carlo results as these parameters vary. While this approach is useful in that it allows us to vary the degree of spatial correlations directly, it has the disadvantage that such simple parameterizations are unlikely to capture the complicated forms of spatial correlations which are likely to be encountered in practice.[9] Instead, we use a data-based method of selecting the matrix

[9] This approach was taken in an earlier draft of the paper. We performed a large number of simulations, allowing the magnitude of spatial correlations and the sizes of the cross-sectional and time-series dimensions to vary. The main findings of this exercise were: 1) OLS with time dummies performs poorly when there is heterogenous response to a common factor and when there are lagged cross-sectional correlations. 2) SUR performs well only when N is very small relative to T and 3) the HAC estimator performs well in all cases, even for values of T as low as 25, and its performance does not depend on the size of the cross-sectional dimension which may be arbitrarily large. The results of these experiments are available from the authors upon request.

of spatial correlations. We estimate an AR(1) process for real output growth for 20 U.S. states[10] and for 24 O.E.C.D. economies, and use the observed variance-covariance matrix of the residuals as Σ to generate spatial correlations in our artificial data.

Table 1 reports the coverage rates for this specification for the cases T=25 and T=50 for each of the three estimators (OLS with time dummies, SUR and HAC). From Table 1, it is clear that both OLS and SUR do quite poorly even when T=50, having coverage rates which never rise above .694 and .679 respectively. Even when SUR is correctly specified (in the first column where $\rho=0$), its performance is quite poor. The reason for this is that even when T=50, the estimates of the contemporaneous spatial correlations matrix which it uses in a FGLS procedure are very imprecise, as it attempts to estimate a very large number of free parameters. In contrast, our non-parametric spatial correlation consistent HAC estimator performs much better than both OLS and SUR, with coverage rates which range from .818 to .903.

[10] Of course, a drawback of this approach is that the estimates of elements of Σ will not be very precise, as the cross-sectional dimension in these regressions is large relative to the time dimension. For this reason, we use only 20 states because annual data for gross state product is only available for 23 years, from 1963-1986. To give some idea of the magnitude of the spatial correlations, note that the average cross-sectional correlation for the U.S. state data is .193, with a maximal value of .629, while for the O.E.C.D., the corresponding figures are .312 and .761.

4 Conclusion

Spatial and other forms of cross-sectional correlation are likely to be an important complicating factor in many empirical studies. We have argued that they are especially likely to arise in macroeconomics and international economics applications in which the cross-sectional units are countries or regions. Standard techniques for dealing with this problem such as the introduction of time dummies or SUR require either restrictive parameterizations of the form of the correlation or pre-estimation of a large number of parameters. In this paper, we have shown that non-parametric covariance matrix estimators of the type proposed by Newey and West (1987) have a simple analog in the panel data case. Asymptotic theory indicates that this technique can accommodate a wide variety of spatial correlations, and moreover, that the size of the cross-sectional dimension is no obstacle to obtaining large-T asymptotic results. This suggests that our technique is applicable to a broad class of empirical studies which look at large cross-sections of countries, states or regions observed over time. Monte Carlo experiments demonstrate that the finite-sample properties of this estimator are good, and are often superior to those of other commonly used techniques.

Finally, we note that this paper has relied exclusively on large-T asymptotics to deliver consistent covariance matrix estimates in the presence of cross-unit correlations. However, when T is small or when there is only a single cross-section, the problem of consistent non-parametric covariance matrix estimation appears much less tractable. The reason for this is that, unlike in the time dimension, there is no natural ordering in the cross-sectional dimension upon which to base mixing restrictions, and hence it is not possible to construct the pure cross-sectional analogs of time-series HAC estimators. Thus, it would appear that consistent covariance matrix estimation in models of a single cross-section with spatial correlations will have to continue to rely on some knowledge of the form of these spatial correlations.

Appendix

Mixing Random Fields

It is most convenient to characterize cross-sectional and temporal dependence in the context of random fields[11]. Let Z^2 denote the two-dimensional lattice of integers, i.e. $Z^2=\{(i,t)|i=1,2,\ldots,N,\ldots,\ t=1,2,\ldots,T,\ldots\}$, and let $(\Omega,\mathcal{F},P)$ denote the standard probability triple. A random field is defined as follows:

Definition: The set of random variables $\{\epsilon_z|z\in Z^2\}$ *on* $(\Omega,\mathcal{F},P)$ *is a random field.*

Next, consider sets of the form $A_t=\{(i,s)|s\leq t\}$. The σ-algebra generated by the collection of random variables whose indices lie in the set A_t, which we denote $\mathcal{F}_{-\infty}^{t}\equiv\sigma(\epsilon_z|z\in A_t)$, has the usual interpretation as the information set available at time t. Furthermore, let $\mathcal{F}_{t+s}^{\infty}\equiv\sigma(\epsilon_z|z\in A_{t+s}^{C})$, where A_t^C denotes the complement of A_t. Using this notation, we can summarize the dependence between two σ-algebras using α-mixing coefficients defined in a manner analogous to the standard univariate α-mixing[12] coefficient, i.e.

$$\alpha(s) \equiv sup_{<t>}\ sup_{<F_1\in\mathcal{F}_{-\infty}^{t},F_2\in\mathcal{F}_{t+s}^{\infty}>} |P[F_1\cap F_2] - P[F_1]P[F_2]|$$

A mixing random field is defined as follows:

Definition: A random field is mixing of size $r/(r-1)$, $r>1$ *if for some* $\lambda>r/(r-1)$, $\alpha(s)=O(s^{-\lambda})$.

---

[11] Random field structures have been developed extensively in the statistics literature. See Rosenblatt (1970), Deo (1975), Bolthausen (1982), and Bulinskii (1988). Some economic applications include Wooldridge and White (1988), Quah (1990), and Conley (1994).

[12] It is straightforward to extend these definitions and the results which follow to ϕ-mixing random fields by defining ϕ-mixing coefficients in the usual way.

This definition of mixing departs from the more standard α-mixing structures on random fields in that it treats the cross-sectional dependence differently from the time-series dependence. Most definitions of mixing[13] restrict the dependence in both dimensions symmetrically, requiring the dependence between two observations to decline as either the distance in the cross-sectional ordering becomes large, or as the time separation becomes large (see, for example, Quah (1990)). This restriction on the dependence across units is required to deliver $(NT)^{1/2}$ asymptotic normality for double sums over i and t of the ϵ_{it}, just as in the one-dimensional case restrictions on the temporal dependence are required to deliver $T^{1/2}$ asymptotic normality for appropriately normalized sums.[14]

The definition of mixing presented here, however, does not restrict the degree of cross-sectional dependence. Instead, we only require the dependence between ϵ_{it} and ϵ_{jt-s} to be small when s is large, for any value of i and j. This is a desirable property, since it will not preclude canonical forms of cross-sectional dependence, such as factor structures in which cross-sectional units may be equicorrelated in a given time period or grouped structures in which observations are correlated according to possibly unobservable group characteristics. This greater permissible cross-sectional dependence comes at the cost that it will not be possible to obtain $(NT)^{1/2}$ asymptotics for double sums over i and t of the ϵ_{it}. However, we do not require this as we rely exclusively on $T^{1/2}$ asymptotics for this double sum.

A useful property of this random field structure is that the sequence of cross-sectional averages of the ϵ_{it} forms a univariate α-mixing sequence, as summarized in the following lemma:.

[13] See Doukhan (1994) for an extensive survey of mixing in random fields and in other contexts.

[14] For such random fields, $(NT)^{1/2}$ asymptotics typically require N and T to go to infinity at the same rate, suggesting that in finite sample applications, the cross-sectional and time-series dimension must be roughly equal for asymptotic approximations to be plausible. For example, Quah (1990) has the restriction that $T=\kappa N$. We do not require this restriction in our asymptotic theory.

Lemma

Suppose that ϵ_{it} is an α-mixing random field of size $r/(r-1)$, $r>1$. Then

$$h_t = \frac{1}{N}\sum_{i=1}^{N} \epsilon_{it}$$

is an α-mixing sequence of the same size as ϵ_{it} for any N.

Proof

The proof is simply a matter of verifying that h_t satisfies the definition of univariate mixing. Define $B_t=\{s|s<t\}\in Z^1$ as the natural one-dimensional analog of A_t, and similarly $\mathcal{G}_{-\infty}^{t}\equiv\sigma(h_z|z\in B_t)$ and $\mathcal{G}_{t+s}^{\infty}\equiv\sigma(h_z|z\in B_{t+s}^{C})$. Define the mixing coefficients for the sequence h_t as $\alpha^h(s)=\sup<t>\sup<G_1\in\mathcal{G}_{-\infty}^{t}, G_2\in\mathcal{G}_{t+s}^{\infty}>|P[G_1\cap G_2]-P[G_1]P[G_2]|$. Now we claim that $\mathcal{G}_{-\infty}^{t}\subseteq\mathcal{F}_{-\infty}^{t}$ and $\mathcal{G}_{t+s}^{\infty}\subseteq\mathcal{F}_{t+s}^{\infty}$. Given this claim, we have $\alpha^h(s)\leq\alpha(s)$ $\forall s$, and hence $\alpha^h(s)$ converges at least as quickly as $\alpha(s)$. Thus the sequence h_t is mixing of the same size of ϵ_{it}.

To verify the claim, note that $h_t:\Omega\rightarrow R^1$ is a Borel function of $\{\epsilon_{it}|i=1,\ldots,N,..\}$, and hence is $\sigma(\epsilon_{it}|i=1,\ldots N,\ldots)$-measurable, i.e. $h_t^{-1}(\mathcal{B})\subset\sigma(\epsilon_{it}|i=1,\ldots,N,..)$ where $\mathcal{B}$ is the σ-algebra generated by the Borel sets. Thus by definition $\sigma(h_t)=\sigma(h_t^{-1}(\mathcal{B}))\subset\sigma(\epsilon_{it}|i=1,..,N,..)$. Finally, note that $\mathcal{G}_{-\infty}^{t}=\sigma(\cup_{s=-\infty}^{t}\sigma(h_s))$ and $\mathcal{F}_{-\infty}^{t}=\sigma(\cup_{s=-\infty}^{t}\sigma(\epsilon_{is}|i=1,..,N,..))$, and so the claim is verified.

This lemma is useful, as it permits us to move from restrictions on temporal and spatial dependence in the random field to simple mixing restrictions on the univariate sequence of cross-sectional averages, h_t.

Proof of Theorem

To prove consistency and asymptotic normality of the GMM estimator, we will verify the conditions in Hamilton (1994), Proposition 14.1. Consistency of the covariance matrix estimator will follow from the arguments of Newey and West (1987). To verify consistency of the GMM estimator (Hamilton, Proposition 14.1, Condition (a)), we need only verify conditions for the consistency of extremum estimators (for example, Amemiya (1985), Theorem 4.1.1). Conditions A and B of Amemiya (1985), Theorem 4.1.1 follow immediately from the compactness of Θ and Assumption 2(a). Condition C of this theorem requires the minimand in the GMM problem to converge uniformly in $\theta\in\Theta$. This condition will be satisfied if the sequence $h_t=h(\theta, z_t)$ obeys a LLN for all $\theta\in\Theta$. Since ψ is a measurable function of z_{it}, it is a mixing random field of the same size as z_{it} by an argument similar to the one use to prove Lemma 1. By Lemma 1, h_t is a univariate α-mixing sequence of size $2(r+\delta)/(r+\delta-1) > r/(r-1)$. Thus, to apply the McLeish (1975) LLN (See White (1984), Theorem 3.47) for α-mixing sequences of size $r/(r-1)$, we need only verify that h_t has finite $(r+\delta)^{th}$ moments. However, to prove consistency of the covariance matrix estimator, we will require the stronger moment condition that $E[|h_t|^{4(r+\delta)}]<\Delta<\infty$. Anticipating this, we verify this condition as follows:

$$
\begin{aligned}
E[|h_t|^{4(r+\delta)}] = E\left[\left|N^{-1}\sum_{i=1}^{N}\psi(\theta,z_{it})\right|^{4(r+\delta)}\right] &\leq N^{-4(r+\delta)}\left[\sum_{i=1}^{N}E[|\psi(\theta,z_{it})|^{4(r+\delta)}]^{1/(4(r+\delta))}\right]^{4(r+\delta)} \\
&\leq N^{-4(r+\delta)}\left[N\Delta^{1/(4(r+\delta))}\right]^{4(r+\delta)} = \Delta
\end{aligned}
\tag{14}
$$

where the first inequality follows from Minkowski's inequality and the second follows from Assumption 2(b). Thus we have verified the conditions for the LLN for the sequence h_t, and the GMM estimator is consistent.

To verify Condition (b) of Hamilton (1994), Proposition 14.1, we need to show that the sequence h_t satisfies a CLT such as White (1984), Theorem 5.19. The CLT require h_t to be a mixing sequence of size $r/(r-1)$ with finite $2r^{th}$ moments. Both these conditions have been verified above. The CLT requires the additional regularity condition that

$$\frac{1}{T}E\left[\left(\sum_{t=a+1}^{a+T} h_t\right)^2\right] \to \bar{\sigma}^2 > 0 \tag{15}$$

uniformly in a as $T\to\infty$. To verify this, observe first that the mixing property of h_t allows us to bound the autocovariances of h_t in the usual way. That is, for s>0, we have $|E[h_t h_{t-s}]| \leq \alpha(s)\Delta$ where $\alpha(s)=O(s^{-(1+\delta)})$ (White, (1984), Corollary 6.16). This corollary requires h_t to be an α-mixing sequence of size $(2+2\eta)/\eta$, $\eta>0$ with $E[|h_t|^{2+2\eta}]<\Delta<\infty$, which may be verified by setting $\eta=(r+\delta-1)$ in the previous paragraph. We can use this bound on the autocovariances of h_t to verify the regularity condition for the CLT, since

$$\begin{aligned}
\frac{1}{T}E\left[\left(\sum_{t=a+1}^{a+T} h_t\right)^2\right] &\leq \frac{1}{T}\sum_{t=a+1}^{a+T} |E[h_t^2]| + \frac{2}{T}\sum_{s=1}^{T-1}\sum_{t=a+1+s}^{T} |E[h_t h_{t-s}]| \\
&\leq \Delta + \frac{2}{T}\sum_{s=1}^{T-1}(T-a-s)\alpha(s) \\
&\leq \Delta + 2\sum_{s=1}^{T-1}\alpha(s) \\
&\leq \Delta'
\end{aligned} \tag{16}$$

The first inequality follows from the triangle inequality and a reorganization of the double sum. The second inequality follows from the previously-derived bounds on the moments h_t and its autocovariances, and the final inequality follows from the summability of $\alpha(s)$. Thus we have verified the required moment and regularity conditions for the CLT. Finally, Condition (c) of Hamilton (1994), Proposition 14.1 is identical to Assumption (3). Therefore, GMM is consistent and asymptotically normal in this random field setting.

To demonstrate consistency of the covariance matrix estimator, we can follow the arguments in Newey and West (1987). Using the definition of the covariance matrix estimator and the triangle inequality, we have

$$
\begin{aligned}
\left|\hat{S}_T - S_T\right| \le & \left|\hat{S}_T - \left[\frac{1}{T}\sum_{t=1}^{T} h_t^2 + \frac{2}{T}\sum_{s=1}^{m} w(s,m)\sum_{t=s+1}^{T} h_t h_{t-s}\right]\right| \\
& + \left|\frac{1}{T}\sum_{t=1}^{T}\left(h_t^2 - E[h_t^2]\right) + \frac{2}{T}\sum_{s=1}^{m} w(s,m)\sum_{t=s+1}^{T}\left(h_t h_{t-s} - E[h_t h_{t-s}]\right)\right| \\
& + \frac{2}{T}\sum_{s=1}^{m}|w(s,m)-1|\sum_{t=s+1}^{T}\left|E[h_t h_{t-s}]\right| \\
& + \frac{2}{T}\sum_{s=m+1}^{T-1}\sum_{t=s+1}^{T}\left|E[h_t h_{t-s}]\right|
\end{aligned}
\tag{17}
$$

To prove consistency, we will show that each of the four terms in this expression tends to zero as $T \rightarrow \infty$. First, however, we derive the following bound, which we will require to show the consistency of the second term in Equation **(17)**:

$$
E\left[\left(\sum_{t=s+1}^{T} Z_{ts}\right)^2\right] \le T\Delta_Z, \quad Z_{ts} = h_t h_{t-s} - E[h_t h_{t-s}]
\tag{18}
$$

The proof of this bound relies on the fact that Z_{ts} is an α-mixing sequence of size $(2+2\eta)/\eta$ with $E[|Z_{ts}|^{2+2\eta}]<\Delta<\infty$. To verify this condition, note that Z_{ts} is a measurable function of a finite number of α-mixing sequences (h_t and h_{t-s}), and hence is α-mixing of the same size as h_t, which is $(2+2\eta)/\eta$ by setting $\eta=(r+\delta-1)$ in Assumption 1. To verify the moment condition, write

$$
\begin{aligned}
E[|Z_{ts}|^{2+2\eta}] &\le \left[E[|h_t|^{2+2\eta}|h_{t-s}|^{2+2\eta}]^{1/(2+2\eta)} + |E[h_t h_{t-s}]|\right]^{2+2\eta} \\
&\le \left[E[|h_t|^{4+4\eta}]^{1/(4+4\eta)} E[|h_{t-s}|^{4+4\eta}]^{1/(4+4\eta)} + E[|h_t^2|]^{1/2} E[|h_{t-s}^2|]^{1/2}\right]^{2+2\eta}
\end{aligned}
\tag{19}
$$

by applying the Minkowski and Cauchy-Schwartz inequalities. However, setting $\eta=(r+\delta-1)$, we have $4+4\eta=4(r+\delta)$, and we have already shown that $E[|h_t|^{4(r+\delta)}]<\Delta<\infty$, and so the moment condition on Z_{ts} is satisfied. The proof of the bound itself is a slight modification of the argument in White (1984), Lemma 6.19, and is similar to the argument used to verify the regularity condition for the CLT above.

Using the bound on the autocovariances of h_t, we can write the fourth term as

$$\frac{2}{T}\sum_{s=m+1}^{T-1}\sum_{t=s+1}^{T}\left|E[h_t h_{t-s}]\right| \le \frac{2}{T}\sum_{s=m+1}^{T-1}\sum_{t=s+1}^{T}\Delta\alpha(s)$$

$$\le \frac{2}{T}\sum_{s=m+1}^{T-1}(T-s)\Delta\alpha(s) \tag{20}$$

$$\le 2\Delta\sum_{s=m+1}^{T-1}\alpha(s)$$

and the final sum will tend to zero as T and m(T) tend to infinity since $\alpha(s)=O(s^{-(1+\delta)})$.

We can again use the bound on the autocovariances of h_t in the third term, resulting in

$$\frac{2}{T}\sum_{s=1}^{m}|w(s,m)-1|\sum_{t=s+1}^{T}\left|E[h_t h_{t-s}]\right| \le 2\sum_{s=1}^{m}|w(s,m)-1|\frac{1}{T}\sum_{t=s+1}^{T}\Delta\alpha(s)$$

$$\le \sum_{s=1}^{m}|w(s,m)-1|\Delta\alpha(s) \tag{21}$$

Since $w(s,m)\to 1$ for all s as $m(T)\to\infty$ and since $\alpha(s)=O(s^{-(1+\delta)})$, this expression tends to zero.

We can follow the argument in Newey and West's Equation (11) to show the consistency of the second term, using a Chebychev's inequality argument and Equation **(18)**.

$$P\left[\left|\frac{1}{T}\sum_{s=1}^{m}w(s,m)\sum_{t=s+1}^{T}Z_{ts}\right| > \epsilon\right] \le P\left[\sum_{s=1}^{m}|w(s,m)|\left|\frac{1}{T}\sum_{t=s+1}^{T}Z_{ts}\right| > \epsilon\right]$$

$$\le \sum_{s=1}^{m}P\left[\left|\frac{1}{T}\sum_{t=s+1}^{T}Z_{ts}\right| > \frac{\epsilon}{Cm}\right]$$

$$\le \sum_{s=1}^{m}\left(\frac{Cm}{\epsilon T}\right)^2 T\Delta_Z \tag{22}$$

$$=\frac{\Delta_Z C^2 m(T)^3}{2\epsilon^2 T}$$

This final term converges to zero by the assumption that $m(T)=o(T^{1/3})$.

Consistency of the first term follows immediately from the consistency of the OLS estimator and the final paragraph in Newey and West (1987).

Spatial Correlations with Factor Structure Representations

The following corrolary to the theorem verifies the claim made in Section 2.3 that it is possible to obtain consistent covariance matrix estimates in the presence of spatial correlations which have a factor structure representation.

Corrolary

Suppose that $y_{it}=x_{it}\beta+\epsilon_{it}$, *with* $\epsilon_{it}=f_t{}'\lambda_i+v_{it}$, $f_t=(f_{1t},\ldots,f_{Mt})'$ *and* $x_{it}=g_t{}'\kappa_i+u_{it}$, $g_t=(g_{1t},\ldots,g_{Pt})'$ *where* λ_i *and* κ_i *are* $Mx1$ *and* $Px1$ *vectors of uniformly bounded constant factor loadings and M and P are finite constants. Suppose further that* $f_{mt}\perp f_{nt}$ *and* $f_{mt}\perp v_{it}$ $\forall m, n, m\neq n$ *and* $\forall t$, *and that* $g_{mt}\perp g_{nt}$ *and* $g_{mt}\perp u_{it}$ $\forall m, n, m\neq n$ *and* $\forall t$, *and that* $E[f_{mt}]=E[g_{mt}]=0$ *and* $E[f_{mt}^2]=E[g_{mt}^2]=1$ $\forall m,t$. *Suppose further that* $\forall i,j,t$ *and m*

(1) (a) $(f_t{}',g_t{}')'$ *is an* α*-mixing sequence of size* $2(r+\delta)/(r+\delta-1)$ *for* $r>1$ *and some* $\delta>0$;

(b) $E[v_{it}]=E[u_{it}]=0$ *and* $v_{it}\perp v_{j,t-s}$ *and* $u_{it}\perp u_{j,t-s}$ *for* $s\neq0$;

(2) (a) $E[x_{it}\epsilon_{it}]=0$;

(b) $E[|f_{mt}x_{it}|^{4(r+\delta)}]<\Delta<\infty$ *and* $E[|u_{it}x_{it}|^{4(r+\delta)}]<\Delta<\infty$;

(3) $E[|g_{mt}|^{2(r+\delta)}]<\Delta<\infty$ *and* $E[|u_{it}|^{2(r+\delta)}]<\Delta<\infty$;

Then the OLS estimator is consistent and asymptotically normal and the panel Newey and West (1987) covariance matrix estimator

$$\hat{V}_T=\frac{1}{Q_T^2}\left(\frac{1}{T}\sum_{j=-m(T)}^{m(T)}w(j,m)\sum_{1<t,t-j<T}\hat{h}_t\hat{h}_{t-j}\right)=\frac{\hat{S}_T}{Q_T^2}$$

where $w(j,m)=1-j/(m(T)+1)$ *and* $m(T)=o(T^{1/3})$, *is a consistent estimator of* V_T *as* $T\rightarrow\infty$ *for any N (including* $N\rightarrow\infty$*).*

Proof

We will prove this corollary by verifying the conditions of the Theorem. To verify Assumption 1, define $\mathscr{F}_{-\infty}^{t} \equiv \sigma((f_s', g_s')' \mid s \leq t)$. Now by a similar argument to the one used in the proof to the Lemma, we have $\sigma((f_s'\lambda_i, g_s'\kappa_i)' \mid (s,i) \in A_t) \subseteq \mathscr{F}_{-\infty}^{t}$ since λ_i and κ_i are uniformly bounded constants. Thus $(f_t'\lambda_i, g_t'\kappa_i)'$ is an α-mixing random field of size $2(r+\delta)/(r+\delta+1)$ since by Assumption 1(a), $(f_t', g_t')'$ is an α-mixing random sequence of size$2(r+\delta)/(r+\delta+1)$. Finally, $(v_{it}, u_{it})'$ is trivially an α-mixing random field of size $2(r+\delta)/(r+\delta+1)$ by Assumption 1(b). Thus we have that $(x_{it}, \epsilon_{it})' = (f_t'\lambda_i, g_t'\kappa_i)' + (v_{it}, u_{it})'$ is an α-mixing random field of size $2(r+\delta)/(r+\delta+1)$.

Assumption 2(a) of the Corollary yields the required orthogonality conditions for the GMM estimator, and Assumption 2(a) of the Theorem is trivially satisfied. The moment condition in Assumption 2(b) of the Theorem follows from the moment conditions in Assumption 2(b) and 3, the assumption that the factor loadings are uniformly bounded, and Minkowski's inequality.

Table 1: Data-Based Monte Carlo Results

Each cell contains coverage rates for OLS with time dummies, SUR and HAC estimators based on 1000 Monte Carlo replications. Contemporaneous cross-unit correlations are generated by computing the cross-correlations of residuals from AR(1) regression of the first difference of Gross State Product for 20 U.S. states from 1963 to 1986 and of Gross Domestic Product for 24 O.E.C.D. economies from 1960 to 1991.

	$\rho=0$	$\rho=.1$	$\rho=.3$	$\rho=.5$
U.S. States	.498	.510	.479	.410
T=25, N=20	.267	.286	.214	.173
	.866	.858	.846	.827
U.S. States	.517	.496	.491	.410
T=50, N=20	.668	.654	.589	.448
	.903	.892	.888	.863
O.E.C.D. Economies	.694	.682	.641	.573
T=25, N=24	.146	.133	.120	.096
	.887	.868	.864	.818
O.E.C.D. Economies	.692	.693	.667	.556
T=50, N=24	.658	.679	.596	.445
	.905	.886	.898	.862

References

Ades, Alberto and Hak B. Chua (1993). "Thy Neighbour's Curse: Regional Instability and Economic Growth". Manuscript, Harvard University.

Al-Najjar, Nabil Ibraheem (1995). "Decomposition and Characterization of Risk with a Continuum of Random Variables". *Econometrica.* Vol 63, No. 5, pp.1195-1224.

Amemiya, Takeshi (1985). *Advanced Econometrics.* Cambridge: Harvard University Press.

Andrews, Donald W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation". *Econometrica.* Vol. 59, No. 4, pp. 817-858.

Bolthausen, E. (1982). "On the Central Limit Theorem for Stationary Mixing Random Fields". *The Annals of Probability.* Vol. 10, No. 4, pp. 1047-1050.

Bulinskii, A.V. (1988). "On Various Conditions for Mixing and Asymptotic Normality of Random Fields". *Soviet Math. Dokl.* Vol. 37, No. 2, pp. 443-448.

Case, Anne C. (1991). "Spatial Patterns in Household Demand". *Econometrica.* Vol. 59, No. 4, pp. 953-965.

Chamberlain, Gary and Michael Rothschild (1983). "Factor Structure and Mean-Variance Analysis on Large Asset Markets". *Econometrica.* Vol. 51, No. 5, pp. 1281-1304.

Conley, Timothy G. (1994). "Econometric Modelling of Cross Sectional Dependence". Working Paper, University of Chicago.

Deo, Chandrakant M. (1975). "A Functional Central Limit Theorem for Stationary Random Fields". *The Annals of Probability.* Vol. 3, No. 4, pp. 708-715.

Doukhan, Paul (1994). *Mixing: Properties and Examples.* New York: Springer-Verlag.

Easterly, William and Ross Levine (1995). "Africa's Growth Tragedy: A Retrospective 1960-89". Manuscript World Bank Policy Research Department.

Elliott, Graham (1993). "Spatial Correlations and Cross-Country Regressions". Manuscript, Harvard University.

Fama, Eugene F. and James D. MacBeth (1973). "Risk, Return, and Equilibrium: Empirical Tests". *Journal of Political Economy.* Vol 81. pp. 607-636.

Froot, Kenneth A. (1989). "Consistent Covariance Matrix Estimation with Cross-Sectional

Dependence and Heteroskedasticity in Financial Data." *Journal of Financial and Quantitative Analysis*. Vol. 24, pp. 333-355.

Hamilton, James D. (1994). *Time Series Analysis.* Princeton: Princeton University Press.

Keane, Michael and David Runkle (1990). "Testing the Rationality of Price Forecasts: Evidence from Panel Data". *American Economic Review* . Vol. 80, No. 4, pp. 715-735.

Kraay, Aart and Jaume Ventura (1995). "Trade and Fluctuations". Manuscript, MIT and the World Bank.

Lehmann, Bruce (1990). "Residual Risk Revisited". *Journal of Econometrics*. Vol. 45, No. 1, pp. 71-97.

McLeish, D. L. (1975). "A Maximal Inequality and Dependent Strong Laws". *The Annals of Probability*. Vol. 3, No. 4, pp. 826-836.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica*. Vol. 55, No. 3, pp. 703-708.

Quah, Danny (1990). "International Patterns of Growth: Persistence in Cross-country Disparities". Manuscript, MIT Department of Economics.

Rosenblatt, M. (1970). "Central Limit Theorem for Stationary Processes". *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. Vol.2, pp.551-561.

White, Halbert (1980). "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity". *Econometrica*. Vol. 48, No. 4, pp. 817-838.

______(1984). *Asymptotic Theory for Econometricians.* New York: Academic Press.

Wooldridge, Jeffrey M. and Halbert White (1988). "Some Invariance Principles and Central Limit Theorems for Dependent Heterogenous Processes". *Econometric Theory*. Vol. 4, No. 2, pp.210-230.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1531	Some New Evidence on Determinants of Foreign Direct Investment in Developing Countries	Harinder Singh Kwang W. Jun	November 1995	S. King-Watson 31047
WPS1532	Regulation and Bank Stability: Canada and the United States, 1870–1980	Michael Bordo	November 1995	D. Evans 38526
WPS1533	Universal Banking and the Financing of Industrial Development	Charles W. Calomiris	November 1995	D. Evans 38526
WPS1534	The Evolution of Central Banking	Forrest Capie	November 1995	D. Evans 38526
WPS1535	Financial History: Lessons of the Past for Reformers of the Present	Gerard Caprio, Jr. Dimitri Vittas	November 1995	D. Evans 38526
WPS1536	Free Banking: The Scottish Experience as a Model for Emerging Economies	Randall Îroszner	November 1995	D. Evans 38526
WPS1537	Before Main Banks: A Selective Historical Overview of Japan's Prewar Financial System	Frank Packer	November 1995	D. Evans 38526
WPS1538	Contingent Liability in Banking: Useful Policy for Developing Countries?	Anthony Saunders Berry Wilson	November 1995	D. Evans 38526
WPS1539	The Rise of Securities Markets: What Can Government Do?	Richard Sylla	November 1995	D. Evans 38526
WPS1540	Thrift Deposit Institutions in Europe and the United States	Dimitri Vittas	November 1995	P. Infante 37642
WPS1541	Deposit Insurance	Eugene White	November 1995	D. Evans 38526
WPS1542	The Development of Industrial Pensions in the United States in the Twentieth Century	Samuel H. Williamson	November 1995	D. Evans 38526
WPS1543	The Combined Incidence of Taxes and Public Expenditures in the Philippines	Shantayanan Devarajan Shaikh I. Hossain	November 1995	C. Bernardo 37699
WPS1544	Economic Performance in Small Open Economies: The Caribbean Experience, 1980–92	F. Desmond McCarthy Giovanni Zanalda	November 1995	M. Divino 33739

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1545	International Commodity Control: Retrospect and Prospect	Christopher L. Gilbert	November 1995	G. Ilogon 33732
WPS1546	Concessions of Busways to the Private Sector: The São Paulo Metropolitan Region Experience	Jorge M. Rebelo Pedro P. Benvenuto	November 1995	A. Turner 30933
WPS1547	Testing the Induced Innovation Hypothesis in South African Agriculture (An Error Correction Approach	Colin Thirtle Robert Townsend Johan van Zyl	November 1995	M. Williams 87297
WPS1548	The Relationship Between Farm Size and Efficiency in South African Agriculture	Johan van Zyl Hans Binswanger Colin Thirtle	November 1995	M. Williams 87297
WPS1549	The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects	Jonathan Isham Daniel Kaufman	November 1995	S. Torres 39012
WPS1550	Governance and Returns on Investment: An Empirical Investigation	Jonathan Isham Daniel Kaufman Lant Pritchett	November 1995	S. Fallon 38009
WPS1551	Sequencing Social Security, Pension, and Insurance Reform	Dimitri Vittas	December 1995	P. Infante 37642
WPS1552	Unemployment Insurance and Duration of Unemployment: Evidence from Slovenia's Transition	Milan Vodopivec	December 1995	J. Walker 37466
WPS1553	Spatial Correlations in Panel Data	John Driscoll Aart Kraay	December 1995	R. Martin 39065